DOCUMENT RESUME

ED 194 631 TM 800 763

AUTHOR Clifford, Paul I.; Matthews, Doris B.
TITLE Public Misconceptions of Tests and How to Overcome Them.
PUB DATE [77]
NOTE 18p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Public Opinion; *Social Attitudes; *Standardized Tests; *Testing Problems

ABSTRACT

Six fundamental public misconceptions about tests are discussed: (1) the view of tests as infallible, objective predictors of behavior versus the view of tests as completely without value; (2) the lack of understanding concerning pyschometric characteristics of tests; (3) the confusion between labels and constructs; (4) the belief that all the test user needs to know about the test is included in the test manual; (5) the belief that testing serves discovery and brokerage purposes in society; and (6) the notion that tests are as effective in assessing performance of the gifted and creative as they are in measuring the behavior of more pedestrian individuals. Conclusions point to the need for test publishers and the professional measurement community to take the lead in developing and using new instruments which are responsive to societal goals in overcoming these misconceptions. (GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED194631

PUBLIC MISCONCEPTIONS OF TESTS AND HOW TO OVERCOME THEM

Paul I. Clifford

Counseling Psychologist

Career Management, Inc.

3384 Peachtree Road, N.E., Suite 930

Atlanta, Georgia 30326

(W) 404-233-5441 (H) 404-799-3805

Doris B. Matthews

Professor of Education

South Carolina State College

Orangeburg, South Carolina 29117

(W) 803-536-7093 (H) 803-536-4936

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
D. Matthews

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM800763

Paul I. Clifford, licensed psychologist, is the former registrar at Atlanta University and the former psychology departmental chairperson at South Carolina State College. Upon partial retirement, he became affiliated with Career Management, Inc., in Atlanta, Georgia. He has been a lecturer on testing and human behavior all over the United States and has written for journals such as the Harvard Educational Review and Human Sexuality.

Doris B. Matthews is an educational psychologist in the Department of Education at South Carolina State College. She is a former counseling supervisor and has written for journals such as The School Counselor and The Employment Service Review and for ERIC. Her research has been accepted for presentation at annual meetings of the American Education Research Association and the American Personnel and Guidance Association.

PUBLIC MISCONCEPTIONS OF TESTS AND HOW TO OVERCOME THEM

Paul I. Clifford

and

Doris B. Matthews

ABSTRACT

Since testing is surrounded by biases, suspicions, claims, and counter-claims, the discussion explores six of the fundamental public misconceptions about them with some suggestions for overcoming the misconceptions. The developmental approach encourages educators, psychologists, and test publishers to join forces in developing and using instruments which measure behavior related to society's goals and "include" rather than "exclude."

PUBLIC MISCONCEPTIONS OF TESTS AND HOW TO OVERCOME THEM

Paul I. Clifford and Doris B. Matthews

Introduction

The use of tests affects not only the lives of practically all individuals in the United States, but also the functioning of society itself. Education at all levels, civil service, the armed services, business and industry, along with various other facets of our national existence are inexorably intertwined with tests and testing.

Although tests are widely utilized and culturally embedded in the fabric of the social order, they are also under constant attack and meet increased resistance on a variety of fronts. Federal truth-in-testing legislation now pending in Congress is an example and, if passed, would vastly reshape the testing industry.

Testing programs in the United States are larger, more routinized and more repetitive than in other countries. The testing programs have developed to the point where they not only influence the personal lives and destinies of countless individuals but they are also a vital factor in shaping America's national characteristics, if not its national destiny. It is for these reasons that tests, which theoretically are neutral tools of no greater philosophical involvement than thermometers, or sphygnometers are abused, misued, misunderstood, resented and attacked.

In some instances, testing has become the sacrificial lamb for much that displeases various sectors of the population, both professional and lay, about American education in particular and American society in general.

It may be that many of the misconceptions about tests are, in part, symptoms which indicate quite validly and faithfully the various underlying imperfections, weaknesses and pathology within the American social order itself. The caveat needs to be raised that treating the distressing symptoms may tend to mask the pathology which the symptoms reflect.

In order to pursue the topic, it is first necessary to present the conceptual framework or set of terms in which the discussion is cast. The term "public" is defined as test developers, test publishers, test users, individuals to whom tests are administered, and individuals within the general population not previously identified. The term "misconceptions" refers to the act or acts of misinterpreting test results, drawing invalid inferences from test data, and using tests inappropriately, irresponsibly, unfairly or in a "biased" manner. The term "tests" in this paper refers to special kinds of assessment procedures which are intended to sample particular aspects of human behavior by requiring the examinee to respond to a set of tasks or questions presented under standardized conditions and designed to yield scores which have the psychometric characteristics of meaningful validity and reliability. "Ways of overcoming misconceptions" refer to suggestions presented in this discussion.

Public Misconceptions of Tests and Suggestions for Overcoming Them

Within the conceptual framework explicated, it is possible now to identify some of the major misconceptions about tests and to present certain suggestions which may contribute to overcoming these negative (and negativistic) notions.

The most fundamental misconception about tests is a polar one. On the one hand there are those who view tests as infallible, completely objective, "scientific" indicators and predictors of individual and group behavior. On the other hand are those who view tests as completely without merit or value. Some people who take this polar view advocate moratoria on testing or total abandonment of tests. The former group tends to hold on to the notion that those aspects of human behavior which have been most frequently subjected to psychological and educational measurement are fixed and immutable through the mechanism of genetic determinism. The latter group tends to view the same aspects of human behavior as consequences of environmental or cultural factors which either facilitate or retard the actualization of potentialities which are residual in all human beings.

The genetic determinists tend to label individuals and groups upon the basis of test results. These labels, either by implication or inuendo or by direct statement, are interpreted in terms of a superiority--inferiority typology. Such labeling is frequently the basis for differential and often prejudicial treatment and opportunities.

Cultural determinists tend to feel and to believe that since no facet of human behavior is unaffected by the culture or subculture into which one is born and in which one is reared, all tests include cultural factors which either help or hinder the individual's test behavior. These persons feel and believe that testing does more harm than good and that the use of tests should be discontinued.

One suggestion for overcoming these polar misconceptions is to view individual behavior as a function of the multiplicative interaction of

genetic, environmental, cultural as well as unknown factors. The measurement of such dynamic, interactional systems requires:

1. Periodic revision whenever changing conditions and/or recent research data demand changes in test instrumentation.
2. Extreme care and high levels of measurement sophistication are needed in order to make valid inferences about legitimate uses and the meaning of test results.
3. Test developers, test publishers, and test users share the responsibility for appropriate, legitimate and responsible use of tests and the interpretation of tests results. Test developers and publishers must take every precaution to obviate the misuse of tests, while test users must assure themselves that their inferences based upon test data are reflective of current research and contemporary theory.
4. Professionals in the testing field must take the responsibility of putting their own house in order and for educating the general public about the appropriate use, as well as the abuse of testing and test results.
5. The testing profession should take the lead in helping test users and the general public distinguish between criterion-referenced, content-referenced, and norm-referenced interpretations of test scores. When a test score is directly related to performance at each point on the achievement continuum being measured, content-referenced interpretations are required. Criterion-referenced interpretations are needed when a test score is directly related to the examinee's performance at any given point on the continuum of an external variable, such as level of

job performance or levels of performance competencies. When it is desirous to compare an individual's test score with normative data, such as centiles or standard score norms, norm-referenced interpretations are called for.

6. When test users decide that norm-referenced interpretations are appropriate, the norms used must refer to precisely defined and clearly described populations. Such populations should be groups with whom test users can legitimately and responsibly compare the person or persons being tested.
7. Extreme care must be taken not to generalize normative test data beyond the specific normative population which provided the norms.
8. Test developers and publishers should warn test users against inappropriate interpretations of test scores.
9. Users of tests should develop local norms for use with specific populations. If prediction is a purpose of testing, expectancy tables or experience tables should be developed by local users to show probabilities of success or proficiency levels with respect to local situations.
10. Test norms must be viewed as the way specific groups perform on tests, not as standards or expectations or goals. Testing is misused when norms become goals. This may be a particularly American disease, for a society that is prone to confuse consensus with values (if everyone is doing it, it must be right--or, at the very least, it must be justifiable) is more prone to confuse norms with standards (if the average child begins to learn how to read at age 6.6, then 6.6 must be the proper age at which a child should begin to learn). Test norms should

be regarded merely as descriptions of the status quo rather than as prescriptions for better education. Just as the lower half of the tested universe must be helped to perform up to the current average, so the upper half of the tested universe must be helped to ignore the current norms entirely, thus pushing the entire distribution toward higher and higher norms. Educational norms are rarely, if ever, natural laws. They exist to be exceeded, to be bettered, to be outdistanced, rather than to be slavishly followed.

11. Terms which are frequently ill-defined and misunderstood, such as "intelligence," "intelligence quotient," "mental age," and "grade-equivalent" should be used sparingly and when used should be precisely defined.

The second set of fundamental misconceptions about tests is related to the basic psychometric characteristics of tests such as objectivity, validity and reliability. Many tests users believe that because tests yield measurement data, such data are largely self-explanatory, precise, and "scientific" thereby eliminating the need for judgment and other subjective interpretations. Many test users have little or no understanding of the basic concepts of validity and reliability.

These misconceptions can be counteracted by careful understanding of the kinds of scales used in educational and psychological measurement and the kinds of assumptions which underlie measurements of human behavior.

Test users need to understand that test validity is inferred, not measured. Validity is inferred from answering three questions: (1) "How well does this instrument serve the purpose with the sample under these

specific conditions?" (2) "What inferences can be drawn about what the specific instrument is measuring within this sample under these conditions?" and (3) "What inferences, if any, can be drawn about other domains of behavior from the test data collected?" Predictive validity, concurrent validity, content validity and construct validity should be inferred holistically. The fact that no test is valid for all purposes, or in all situations, or for all groups, or for all individuals, should be impressed upon test users.

Reliability refers to the degree to which a test user can attribute test results to systematic sources of variation. This means that different means of estimating reliability must take into account different sources of variation or error. Test users must be able to estimate the precisely defined components of score variance in order to interpret test scores with optimal effectiveness.

Test users should become more adept at using the standard error of measurement which is an estimate of the root-mean-square departure of a series of observed scores from their corresponding true scores. Standard errors of measurement are more appropriate means of interpreting test scores than are reliability coefficients which are used for inter-test comparisons or contrasts.

A third misconception about tests has to do with labeling and the confusion between labels and constructs. Individuals often think that by giving a kind of behavior a name or label one has also gone some distance in understanding or explaining that behavior. For example, one might label a child "mentally retarded" or "culturally disadvantaged" as the result of

some kind of test performance. After the label is applied, educators then begin to "explain" the behavior and even prescribe certain "treatments" for the "condition" indicated by the label.

Constructs are not labels and labels are not constructs. A construct is an idea which has been developed or constructed through scientific imagination. Constructs are theoretical ideas developed to explain or to organize some dimension or aspect of existing knowledge. A label, on the other hand, is simply an identifying or descriptive name, word or phrase. In testing, individuals appear to be more enamored with labels than with constructs, but more devastatingly educators are prone to reify which is "explanation" by labeling. This misconception could be obviated by emphasizing constructs and by cessation of labeling.

Another misconception about tests is that all the test user has to know about the test is that information presented in the test manual. This misconception can be eliminated only if the use of tests in schools, business, industry, or in any other situation is restricted to those who have:

1. General and specific contemporary knowledge of the principles of the measurement of human behavior and the legitimate interpretation and use of such data.
2. A clear referent for the term "bias" which enables one to recognize and avoid bias in test selection, administration, and interpretation.
3. Clearly delineated and explicated their implicit and explicit hypotheses, assumptions, values and purposes concerning test development, use and interpretation.

4. An awareness and appreciation of their own experience and sophistication in measurement and the likely effects of past experiences upon examinees.
5. Worked out a logical and psychological rationale for the use or non-use of cutting scores.
6. Skill in interpreting a given test score as an estimate of performance under a specific set of conditions rather than presenting score interpretations as absolute, immutable characteristics of the examinee which are widely generalizable to other situations, conditions, and circumstances.
7. Developed skill in reporting and interpreting test scores to examinees, parents or third parties. This is particularly important in view of "The Family Educational Rights and Privacy Act of 1974" which gives parents the legal right to see all official records, files, and data directly related to their children from the pre-school years through post-secondary education. Such data specifically include standardized achievement test scores, scores on standardized mental ability, aptitude and other psychological tests, and interest inventories.

Many critics admit that testing serves some useful "discovery" and "brokerage" purposes in the always complex and sometimes chaotic world of American society. Nevertheless, this does not lead them to conclude that even these purposes are equally well performed for all population groups or that these purposes are of equally great importance in maximizing the educational potential of all population groups in current American society. The testing instruments utilized are, decidely, tests of developed ability

and of attained potential. Only to the extent that all attainment is necessarily a reflection of potential for attainment are these tests indirectly indicative not only of what is manifest but also of what is latent in the ability of individuals or groups. This indirect reflection of intellectual potential may be an acceptable index of actual potential for children, youth, and adults who have been substantially encouraged, or at least permitted, to develop in accord with their basic intellectual potential. In such cases--and in society the number, fortunately, is very great--achievement may even be the first and most parsimonious indicator available to us of basic potential for intellectual achievement.

However, when one turns to the culturally and socially different persons, the situation is far different. Here, potential is largely unrecognized, unawakened, and underdeveloped. Here intellectual achievement is low regardless of basic potential and, therefore, developed ability and attained potential can no longer be considered as good indices of basic intellectual potential. Here the testing programs cannot claim to unearth the talent that the social order has suppressed. Actually, the testing programs are geared to recognize talent and potential that are already apparent. They largely confirm talent that the school has already recognized, that the family has already noted, and to which school, family, and community have already contributed much of the facilitative attention currently at their command. However, this deficiency is probably not inherent in testing as such. Whenever it is encountered in testing programs, educators must recognize it as indicative of a deficiency in the society which sponsors these programs, a deficiency which can lead one and

the same society to search madly for the talents that have already been found while ignoring the talents that need careful prospecting if they are to be found at all. With proper societal orientation, testing programs might be devised that would search for such talents as well.

Society's primary task in conjunction with the culturally and socially different is one of development, provided such persons opt for such development, through new methods and materials, rather than one of search and identification through currently available geiger counters. Verbal abilities must be developed, beginning with the lowest levels of labeling. Listening abilities must be liberated and trained. Cognitive styles must be changed or balanced in the direction of academic symbols. Enculturation into a new system of values, motives, and rewards--into a new life style--must be accomplished. Far greater and far more technical teaching expertise must be attained and far better gradation and organization of teaching materials. Cognitive maps must be extended and neighborhood as well as home environments must be guided if not manipulated. Current testing instruments can make very little contribution toward any of these developmental goals.

A final misconception about tests is the notion that tests are as effective in assessing the performances of the gifted, the creative, and non-conformists as they are in measuring the behavior of more pedestrian groups and individuals. That testing is a child of the post-industrial age is undeniable. There is still a question, however, whether this child of the post-industrial age does not itself help spawn mechanical men. What of the mind whose brilliance does not run in the groove hewed by

multiple-choice questions, computer technology, formal psychometrics and one, two, or three hour examinations? There is a deep-seated humanist protest against the testing movement. This protest begins by suspecting that the validity coefficients, the delta indices, the Flanagan r's, and all of the technical aramentarium of the testing trade are no more than the modern day counterpart of the "Emperor's New Clothes" since it is impossible to gauge the intellect from ability to perform essentially trivial, uncreative, eurekaless, and anti-intellectual stunts. And it ends by implying that testing programs, successful as they may be in identifying the more pedestrian varieties of intellect, completely fall short of the mark when it comes to identifying the free spirits, the brilliant nonconformer, the untrammeled minds and the divergent, creative and innovative thinkers. Educators can overcome this misconception if they continue research and development in the assessment of behavior which transcends the mundane.

Conclusion

The discussion has dealt with six fundamental public misconceptions about tests and has presented some suggestions for overcoming these misconceptions. It was pointed out that testing is surrounded by biases, suspicions, claims, counter-claims, and misconceptions. This is a period in American society when numerous and strenuous efforts are being made to break with a past which has not been concerned with the actualization of the potentialities of the individual human being; however, there is little evidence to lend any credence to the belief that the future will be anything other than an elaboration and proliferation of the past. If

individuals seriously and constantly re-evaluate, re-define, and refine societal goals, ends, purposes, and personal values, educators can then determine whether tests foster or harm the goals professed. Test publishers, testing firms, and the professional measurement community must take the lead in developing and using new instruments which are responsive to societal goals and in overcoming the misconceptions which vitiate the current instruments. If educators do not rise to the occasion, the vacuum which exists will be filled with something else which is more likely to be more than less sinister in contrast to what now is available.

References

American Personnel and Guidance Association. The responsible use of tests: a position paper of AMEG: APGA: and MCME. Measurement and Evaluation in Guidance, 1972, 5, 385-388.

American Psychological Association, Inc. Standards for educational and psychological tests. Washington, D.C.: 1971.

Fishman, J. A. and Clifford, P. I. What can mass-testing programs do for-and-to the pursuit of excellence in American education? Harvard Educational Review, 1974, 34, 63-79.

McKenna, B. What's wrong with standardized testing? Today's Education, 1977, 54, 36.

Rudman, H. C. The standardized test flap. Phi Delta Kappan, 1977, 59 (3), 179-185.

Silberman, A. The tests that cheat our children. McCall Magazine, 1977, 64, 214-222.

Tracterberg, P. and Jacoby, E. Pupil testing: a legal view. Phi Delta Kappan, 1977, 59 (4), 249-254.

Womer, F. B. Tests: misconceptions, misuse and overuse. Educational and Psychological Measurement Contributions to Theory and Practice. Edited by David A. Payne and Robert F. McMorris. Waltham, Mass.: Blaisdell Publishing Company, 1977, 229-236.